DOCUMENT RESUME

ED 451 309 UD 034 090

AUTHOR McQuillan, Jeff
TITLE The "Poor Quality" of Bilingual Education Research: Compared to What?
PUB DATE 2000-04-00
NOTE 11p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000).
PUB TYPE Opinion Papers (120) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Bilingual Education; *Educational Research; Elementary Secondary Education; Research Methodology
IDENTIFIERS *Research Quality

ABSTRACT

Critics of bilingual education claim that research supporting native language instruction is weak, a claim that has been echoed by some prominent supporters of bilingual education. This claim has had a damaging effect on the political fate of bilingual education in some states. This paper argues that the primary metric used to support this critique, the percentage of research studies meta-analysts consider methodologically acceptable, is a vague and not widely-accepted approach for weighing the quality of research. Data for the study come from prominent research reviews in the field of education and social sciences and from a random sample of empirical literature reviews from two major journals of research reviews. This paper suggests that the percentage of studies found methodologically acceptable in bilingual education research is not very different from similar federally funded research in education and the social sciences. It notes that there is little basis for comparison for bilingual education research and other psychology- and education-related literatures, since percentages of methodologically acceptable studies are rarely reported in research reviews. It concludes that higher quality research is necessary but should not be viewed in isolation to real-world constraints on such endeavors. (Contains 10 references.) (SM)

Reproductions supplied by EDRS are the best that can be made
from the original document.

UD

ED 451 309

The "Poor Quality" of Bilingual Education Research: Compared to What?

Jeff McQuillan, Assistant Professor

College of Education, Arizona State University

Email: mcquillan@asu.edu

Paper presented at the Annual Meeting of the

American Educational Research Association

New Orleans, LA

April, 2000

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Running Head: POOR QUALITY

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

J. McQuillan

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

BEST COPY AVAILABLE

UD034090

The "Poor Quality" of Bilingual Education Research: Compared To What?

Introduction

Political and academic critics of bilingual education have claimed that the research evidence supporting native language instruction is weak, a claim that has also been echoed by some prominent supporters of bilingual programs. This claim has had a damaging effect on the political fate of bilingual education in states such as California. In this paper, I will argue that (1) the primary metric used to support this critique--the percentage of research studies meta-analysts have found "methodologically acceptable"--is a vague and not widely accepted approach for weighing the quality of a research literature; (2) the percentage of studies found methodologically acceptable in bilingual education research is not very different from similar other, similar federally-funded research in education and the social sciences; (3) there is little basis for comparison for bilingual education research and other psychology- and education-related literatures, since percentages of methodologically acceptable studies are rarely reported in research reviews; and (4) higher quality research is necessary, but should not be viewed in isolation to real-world constraints on such endeavors.

Background

A National Research Council report (August & Hakuta, 1997) recently argued that using program evaluation to determine which type of program is "best" for language minority children has "little value" given the complexities of the components involved (p. 149). Nevertheless, the political debate surrounding bilingual education has focused heavily on the effectiveness issue. The results of program evaluation research have had a significant impact on the public rhetoric surrounding bilingual programs in states with large language minority populations such as

BEST COPY AVAILABLE

California (McQuillan & Tse, 1996; Crawford, 1999). Commenting on the quality of that evaluation research, vocal opponents have stated that it is "worthless" (Rodriguez, 1997, cited in Crawford, 1999). Even supporters of bilingual education have commented that there is a "disappointing percentage of studies...[found] to be methodologically adequate," and have lamented the poor quality of the research in the field (August & Hakuta, 1997, p. 146). These pronouncements have had deleterious effects on the press coverage and editorial commentary on bilingual education (McQuillan & Tse, 1996), and very likely on the outcomes of a recent anti-bilingual education initiative (Crawford, 1999).

Program evaluators have extensively discussed the issues surrounding bilingual program evaluations in the past, noting that they are fraught with difficulties of research design and analysis (Willig & Ramirez, 1993). Lam (1992), for example, noted that there are several problems inherent in bilingual program evaluations, among them: high attrition rates, differing cultural backgrounds of English Language Learners, a limited number of psychometrically acceptable oral language proficiency instruments, and--most critically--difficulty in creating control groups that are truly comparable. These problems are compounded by conflicting federal and state policies, a historically inefficient means of disseminating appropriate evaluation assistance to program evaluators, and a lack of experience by many local evaluators in appropriate research design, the theory and practice of bilingual education, or both.

In spite of--or perhaps because of--these difficulties, Lam and others have reported that bilingual education research is of low quality, a judgment based in part upon the percentage of studies found to be methodologically "acceptable" among evaluations examined by meta-analysts. Lam reports the mean number of acceptable studies found by various meta-analysts from the early 1970s up through the late 1980s was only around 10%. While later reviews (e.g.

BEST COPY AVAILABLE

Rossell & Baker, 1996) found as many as 24% of the studies they reviewed to be acceptable, this is, for prominent opponents as well as some proponents, still considered to be an indicator of poor quality.

Lam (1992) pointed out that bilingual education is not alone in problems related to evaluation. Evaluation efforts in the 1970s and '80s were considered of generally low quality across many areas of education, including special education, migrant education, compensatory education, school desegregation, and others. While this does not, as Lam states, "excuse bilingual educators... from responsibilities for deficiencies in their program evaluations" (p. 183), it does give us a more appropriate context with which to make more balanced judgments about the quality of that research.

Research Issues

This paper critiques the claim that bilingual education evaluation research is of generally low quality based upon the "percentage acceptable" metric. That metric is determined by calculating the proportion of studies deemed methodologically acceptable according to the (varying) criteria of meta-analysts to the total number of studies located on a topic. This critique is carried out in two ways:

1. An examination of the logic of using the "percentage acceptable" metric for evaluating the quality of a research literature;
2. An examination of other areas of meta-analytical research in education and psychology in order to provide a context for the assessment of bilingual education research quality.

Data Sources

The data sources for this study are:

(a) several prominent research reviews in the field of education and the social sciences, including Head Start (GAO, 1997), other federally-funded social science research projects (Cook & Gruder, 1977), and early reading (Stahl & Miller, 1989); and

(b) a random sample of empirical literature reviews that appeared in two major journals of research reviews, the *Review of Educational Research* (N = 11) and *Psychological Bulletin* (N = 16) for the years 1995-996. Only those reviews that included some statistical or vote-count method of comparing treatments or conditions were included.

These two sources will provide a context within which to examine the claim that, by some common standard of research practice, the quality of bilingual education research is "low."

Methods

The analysis was based both on an examination of the logic of the measure in question--that is, is the "percentage acceptable" method a good way to judge research quality?--and a comparison to similar research reviews in other areas of education and psychology, based upon the data sources listed above. These reviews were read to determine the percentage of "acceptable studies" found, and those percentages (if reported) were compared to those found in a recent (critical) review of the bilingual education literature.

Results

Results are presented in two sections:

1. The Logic of the Metric: There are several problems inherent in the "percentage acceptable" metric chosen to judge the quality of bilingual education research. First, the percentage of studies obtained will clearly depend on how many studies are gathered and

BEST COPY AVAILABLE

inspected. This number has varied widely from review to review, from more than 1,400 to less than 20. Okada et al. (1982, cited in Lam, 1992), for example, found 168 studies that were methodologically acceptable, yet this represented only 12% of the total population of studies examined (1,411). Put another way, a "low" percentage might be more than adequate for the purposes of providing an evidential basis for a given educational practice, if the absolute number of studies is high. Clearly it is better to have 10% acceptability of 1000 studies than 90% acceptability of 10 studies. The real question is: Are there a sufficient number of studies to support bilingual education, especially in comparison with the resources devoted to it? Even the most severe critics of bilingual education found 72 acceptable studies (Rossell & Baker, 1996), a number which is greater than the total number of studies considered in other research reviews for areas of education with far wider impact and expense (e.g. Stahl & Miller's (1989) review of early reading approaches, which examined 51 studies). Of course, much of the variation in absolute numbers will depend on the method of research review (vote-count vs. calculation of effect sizes), and on differing exclusion criteria. This is precisely the point: these determinations are rarely uniform across or even within fields of research. Second, many of the "studies" that are included in the research reviews are mandated program evaluations, products of the Title VII regulations for federally-funded programs. As such, they are neither considered part of the published literature nor subject to even the most minimal review by other researchers. They would most likely never be part of a pool of reviewed research in most other areas of education or psychology. These evaluations are written by school district or outside evaluators, many of whom lack essential knowledge of either research design or bilingual education (Lam, 1992). As such, it is not surprising that the percentage will be low, given the pool of "research" that is examined. In Rossell and Baker's review of 300 studies, the vast majority (89%) of those found

BEST COPY AVAILABLE

to be "methodologically unacceptable" consisted precisely of such unpublished studies. Third, excluding studies based on *a priori* research design conditions itself violates an important recommendation made by prominent meta-analysts. They suggest including all studies with sufficient data in a research review in order to determine whether and which research flaws are related to outcomes.

2. Comparisons to Other Meta-Analyses: Two types of comparisons were made with other areas of education and psychology in assessing the appropriateness of the "percentage acceptable" method. First, other reviews of federally-funded projects were examined to see how Title VII and other bilingual program evaluations compared in terms of their quality. Very little data on the "percentage acceptable" for other types of research were found, but those that were located were strikingly similar to bilingual program evaluations. The General Accounting Office's review of 200 Head Start evaluations determined that approximately 10% (22) met their methodological criteria, criteria that were much less stringent than those used by bilingual education reviewers such as Rossell and Baker (1996). Cook and Gruder (1978) found that the percentage of acceptable federal-funded program evaluations contemporaneous with the majority of bilingual evaluations (pre-1980) was in the 10-15% range, again, similar to the results reported by Lam. Other prominently cited meta-analyses in education either made no mention of the number of studies rejected for methodological reasons, or had percentages in a similar range. Stahl and Miller (1989), for example, determined that only nine of the 51 (17.6%) studies they reviewed on early reading methods met one of Rossell and Baker's key criteria for quality--controlling for initial group differences.

A second comparison was made by reviewing systematically meta-analyses in education and psychology found in two major journals of research reviews, the *Review of Educational*

BEST COPY AVAILABLE

Research and *Psychological Bulletin*. Of the 26 randomly selected empirical reviews, only one reported the number of studies that appeared to be rejected explicitly for methodological quality (versus other possible exclusion criteria, such as not examining the constructs or population of interest). That study (Greenwald, Hedges, & Laine, 1996, on school finance) found that 18% of the 175 articles and books initially reviewed met all of their criteria for inclusion, a figure not much different from the mean acceptable percentage from bilingual education reviews through 1996 (15%, range: 5-44%). The absolute number of studies used in the meta-analyses ranged widely (education: 26-133; psychology: 14-286), with a mean number of studies close to those used by more recent bilingual education reviews (education: 55.35 (SD: 30.1); psychology: 90.56 (SD: 73.67); bilingual education: 72 in Rossell & Baker (1996)). These comparisons are quite favorable to bilingual education, especially when one considers that research design difficulties such as the establishment of a comparable control group are much less severe in other areas of education and psychology than they are for bilingual program evaluation.

Conclusion

Educational practice is build upon an imperfect evidential base, as is the case for all social sciences. The field of bilingual education needs better-designed and implemented research studies, as August and Hakuta and others have concluded. But this is quite different from judging the quality of the extant group of studies to be somehow below a standard used for other educational research, however "awful" that may appear to some (e.g. Kaestle 1993). There is no logical or empirical basis for the harsh assessments that have been made of bilingual education evaluations. The "percentage acceptable" method used by other reviewers has little acceptance in either education or psychology as a metric of quality, and is in any case an unstable product of shifting acceptability criteria, with little regard for the absolute number of

BEST COPY AVAILABLE

studies available. It is, in other words, the sort of crude and context-free "single statistic" which as research methodologists have warned as constituting the poorest way to make a reasoned argument (Abelson, 1996). Future evaluations of bilingual education research quality need to take into account the broader context of educational evaluation in general, and the not unfavorable position that bilingual evaluation holds in that context.

References

Abelson, R. (1995). *Statistics as principled argument*. Hillsdale, NJ: Lawrence Erlbaum Associates.

August, D., & Hakuta, K (Eds.). (1997). *Improving schooling for language minority children: A research agenda*. Washington, D.C.: National Academy Press.

Cook, T., & Gruder, C. (1978). Metaevaluation research. *Evaluation Quarterly, 2*, 5-51.

Crawford, J. (1999). The campaign against Proposition 227: A post mortem. *Bilingual Research Journal, 21*(1), 1-30.

General Accounting Office. (1997). *Head Start: Research provides little information on impact of current program. Report to the Chairman, Committee on the Budget, House of Representatives*. Washington, D.C.: Author.

Kaestle, C. (1993, January/February). The awful reputation of education research. *Educational Researcher*, 23-31.

Lam, T.C.M. (1992). Review of practices and problems in the evaluation of bilingual education. *Review of Educational Research, 62*, 181-203.

McQuillan, J., & Tse, L. (1996). Does research matter? An analysis of media opinion on bilingual education, 1984-1994. *Bilingual Research Journal, 20*, 1-27.

Rossell, C., & Baker, K. (1996). The educational effectivness of bilingual education. *Research in the Teaching of English, 30*, 7-73.

Willig, A., & Ramirez, J.D. (1993). The evaluation of bilingual education. In M.B. Arias & U. Casanova (Eds.)., *Bilingual education: Politics, practice, research*, (pp. 65-87). Chicago, IL: University of Chicago Press.